The Informational Approach to Global Optimization in presence of very noisy evaluation results. Application to the optimization of renewable energy integration strategies

Héloïse Dutrieux[1,3] & Ivana Aleksovska[2] & Julien Bect[2]
& Emmanuel Vazquez[2] & Gauthier Delille[1] & Bruno François[3]

[1] *Département Économie, Fonctionnement et Études des Systèmes Energétiques (EFESE), EDF R&D, Clamart, France /* prenom.nom@edf.fr

[2] *Laboratoire des Signaux et Systèmes (L2S, UMR CNRS 8506) CentraleSupélec, CNRS, Université Paris-Sud, Gif-sur-Yvette, France*
prenom.nom@centralesupelec.fr

[3] *Laboratoire d'Électrotechnique et d'Électronique de Puissance de Lille (L2EP) École Centrale de Lille, Villeneuve d'Ascq, France /* prenom.nom@ec-lille.fr

Résumé. Nous considérons le problème de l'optimisation globale d'une fonction f à partir d'évaluations très bruitées. Nous adoptons un point de vue bayésien séquentiel : les points d'évaluation sont choisis de manière à réduire l'incertitude sur la position de l'optimum global de f, cette incertitude étant mesurée par l'entropie de la variable aléatoire correspondante (Informational Approach to Global Optimization, Villemonteix et al., 2009). Lorsque les évaluations sont très bruitées, l'erreur d'estimation de l'entropie par simulation conditionnelle devient non négligeable par rapport à ses variations sur son domaine de définition. Nous proposons une solution à ce problème en choisissant les points d'évaluation comme si plusieurs évaluations allaient être faites en ces points. Une application à l'optimisation d'une stratégie d'insertion des énergies renouvelables dans un réseau de distribution d'électricité illustre la méthode proposée.

Mots-clés. Processus gaussiens ; Planification et analyse d'expériences numériques ; Optimisation bayésienne ; Énergies renouvelables ; Réseau de distribution électrique.

Abstract. We consider the problem of global optimization of a function f from very noisy evaluations. We adopt a Bayesian sequential approach: evaluation points are chosen so as to reduce the uncertainty about the position of the global optimum of f, as measured by the entropy of the corresponding random variable (Informational Approach to Global Optimization, Villemonteix et al., 2009). When evaluations are very noisy, the error coming from the estimation of the entropy using conditional simulations becomes non negligible compared to its variations on the input domain. We propose a solution to this problem by choosing evaluation points as if several evaluations were going to be made at these points. The method is applied to the optimization of a strategy for the integration of renewable energies into an electrical distribution network.

Keywords. Gaussian processes; Design and Analysis of Computer Experiments; Bayesian Optimization; Renewable Energies; Electrical Distribution Network.

1 Introduction

Let f be a continuous real-valued function, defined on $\mathbb{R}^d$ (or a subset of $\mathbb{R}^d$), $d \geq 1$. Given a finite set $\mathbb{X} \subset \mathbb{R}^d$, we consider the problem of estimating the minimum $M = \min_{x \in \mathbb{X}} f(x)$ and the corresponding set of minimizers, $x^\star \in \operatorname{argmin}_{x \in \mathbb{X}} f(x)$, using a sequence of evaluations of f at points $X_1, X_2, \ldots X_n \in \mathbb{X}$. In this article, the evaluation results are assumed noisy: at each X_i, we observe a perturbed value of $f(X_i)$. The construction of an optimization algorithm $\underline{X} = (X_1, X_2, \ldots)$ is viewed as a sequential decision problem: given n (noisy) evaluation results at $X_1, \ldots, X_n$, we must choose X_{n+1} in order to get, in the end, the best estimators of $x^\star$ and M according to a certain loss function.

We adopt the following (classical) Bayesian approach for constructing $\underline{X}$. The unknown function f is considered as a sample path of a Gaussian random process ξ defined on some probability space $(\Omega, \mathcal{B}, \mathsf{P}_0)$, with parameter $x \in \mathbb{X}$. Then, a noisy evaluation of f at $X_i \in \mathbb{X}$ is modeled by the random variable $\xi_i^{\text{obs}} := \xi(X_i) + \varepsilon_i$, $i = 1, 2, \ldots$, with $\varepsilon_1, \varepsilon_2 \ldots \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2)$ (here, σ^2 is assumed to be known). Denote by P_n the conditional distribution $\mathsf{P}_0(\cdot \mid \mathcal{I}_n)$, where $\mathcal{I}_n = \left\{X_1, \xi_1^{\text{obs}} \ldots, X_n, \xi_n^{\text{obs}}\right\}$, and by E_n and var_n the conditional expectation $\mathsf{E}(\cdot \mid \mathcal{I}_n)$ and conditional variance $\text{var}(\cdot \mid \mathcal{I}_n)$ respectively. Following Villemonteix et al. (2009) and Vazquez et al. (2008), the efficiency of an algorithm $\underline{X}$ after n evaluations is measured using the posterior Shannon entropy

$$H(x^\star; \mathcal{I}_n) = -\sum_{x \in \mathbb{X}} \mathsf{P}_n(x^\star = x) \log \mathsf{P}_n(x^\star = x)\,, \tag{1}$$

which quantifies the residual uncertainty about the position of $x^\star$. Then, each new evaluation point is chosen using a *Stepwise Uncertainty Reduction* (SUR) approach, which consists in minimizing a sampling criterion J_n that corresponds to the expected residual uncertainty on $x^\star$ after $n+1$ evaluation results:

$$X_{n+1} = \operatorname{argmin}_{x \in \mathbb{X}} J_n(x) \quad \text{with} \quad J_n(x) := \mathsf{E}_n\left(H(x^\star; \mathcal{I}_{n+1}) \mid X_{n+1} = x\right). \tag{2}$$

Notice that $J_n(x)$ is an expectation with respect to the random evaluation result ξ_{n+1}^{obs} at $X_{n+1} = x$. Minimizing J_n is equivalent to maximizing the mutual information between $x^\star$ and ξ_{n+1}^{obs}. The reader is referred to Picheny et al. (2013) to a review of other sampling criteria for noisy optimization.

From a numerical point of view, the computation of J_n is based on two approximations. A first approximation is required for the computation of the expectation in (2) with respect to the posterior distribution of ξ_{n+1}^{obs} at $X_{n+1} = x$. Since ξ and the evaluation noise are Gaussian, the expectation in (2) is a one-dimensional integral with respect to the Gaussian posterior density of ξ_{n+1}^{obs}, which can be carried out with a standard Gauss-Hermite quadrature. A second approximation is needed to compute the entropy of the posterior distribution of $x^\star$. Villemonteix et al. (2009) estimate this entropy by plugging

into (1) an estimator of $\mathsf{P}_n(x^\star = x)$, with x ranging over $\mathbb{X}$, which, in turn, is estimated by Monte-Carlo simulations of sample paths of ξ conditioned on $\mathcal{I}_n$.

When evaluations are noise-free, it is often possible to obtain a satisfactory estimator of the entropy with a moderately large number of sample paths (≈ 1000). However, when the evaluation noise becomes large, it appears that, for the same moderately large number of sample paths, the variance of estimation of the entropy becomes non negligible with respect to the information provided by a *single* evaluation. Then, minimizing J_n to choose new evaluation points becomes questionable. In this article, we propose to circumvent this problem with a new sampling criterion where, in essence, we pretend that several evaluations are going to be carried out instead of a single one.

2 The Informational Approach to Global Optimization with (very) noisy evaluations

Since a single noisy evaluation provides limited information about $x^\star$, and therefore yields by itself little progress in the optimization procedure, the variations of J_n on $\mathbb{X}$ can be dominated by its estimation error (as illustrated in Figure 1, first left).

A natural idea to gain more information from noisy evaluations is to perform several evaluations at each iteration of the optimization algorithm. Our contribution is as follows: we suggest to build a sampling criterion J'_n such that for all $x \in \mathbb{X}$, $J'_n(x)$ corresponds to the expected residual uncertainty about $x^\star$ when K (noisy) evaluations of f are performed at x:

$$J'_n(x) := \mathsf{E}_n \left(H(x^\star; \mathcal{I}_{n+K}) \mid X_{n+1} = \ldots = X_{n+K} = x \right). \tag{3}$$

The resulting criterion is illustrated in Figure 1 with K equal to 10, 100 and $+\infty$.

We refer to K as the *virtual batch size*, since we do not actually intend to perform K evaluations at the minimizer X_{n+1} of J'_n. Once X_{n+1} has been obtained by minimizing (3), any number K_0 of evaluations (between $K_0 = 1$, as assumed in Section 1, and $K_0 = +\infty$) can actually be performed at this point; this number K_0 is the *actual batch size*. We suggest to take K large enough to make the error of estimation of J'_n small with respect to the variations of the criterion, and to carry out only one actual evaluation ($K_0 = 1$) at each iteration if evaluations are very expensive, or a batch of size $K_0 > 1$ (typically, $K_0 \ll K$) if evaluations are only moderately expensive or if parallel processing is available. Another possibility would be to update K at each iteration so as to consider the whole remaining budget of evaluations as suggested in Picheny et al. (2010).

The idea of considering K evaluations *at the same point* in (3) is only an artificial construction, motivated by the fact that the numerical complexity of the computation of J'_n is the same as that of J_n. Indeed, it can be shown that the distribution of ξ conditioned on $\mathcal{I}_{n+K}$ only depends in this case on $\xi_1^{\text{obs}}, \ldots, \xi_n^{\text{obs}}$ and $\bar{\xi}_{n+1} = \frac{1}{K} \sum_{k=1}^{K} \xi_{n+k}^{\text{obs}} = \xi(x) + \frac{1}{K} \sum_k \varepsilon_{n+k}$. This has two consequences. First, the expectation in (3) is simply

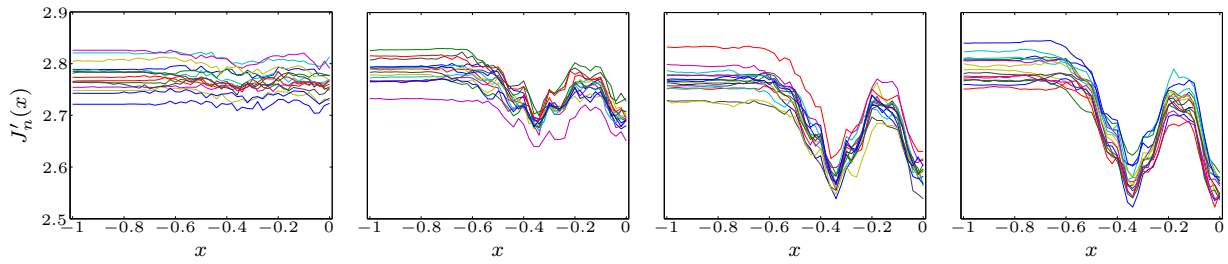

Figure 1: Realizations of the numerical estimate of the sampling criterion $J'_n(x)$ for the data shown in Figure 2 (right). Each figure represents 15 independent realizations (corresponding to independent samples of conditional simulations). The batch size is, from left to right: $K = 1, 10, 100$ and $+\infty$. A standard 15-order Gauss-Hermite is used for the integration and 1000 conditional samplepaths.

a one-dimensional integral with respect to the (conditional) distribution of $\bar{\xi}_{n+1}$, which is Gaussian, with mean equal to $\mathsf{E}_n(\xi(x))$ and variance equal to $\mathrm{var}_n(\xi(x)) + \frac{1}{K}\sigma^2$. Second, the simulation of sample paths of ξ conditioned on the $n + K$ random variables $\xi_1^{\mathrm{obs}}, \ldots, \xi_{n+K}^{\mathrm{obs}}$ boils down to the simulation of sample paths of ξ conditioned on the $n+1$ random variables $\xi_1^{\mathrm{obs}}, \ldots, \xi_n^{\mathrm{obs}}, \bar{\xi}_{n+1}$.

The optimization algorithm with the new criterion J'_n is available for testing in a development branch of the STK toolbox (Bect et al., 2014).

3 Application

The method is applied to the optimization of a strategy for the integration of Renewable Energy Sources (RES) into an electrical distribution network. This strategy describes how the Distribution System Operator (DSO) connects new producers to the network under strict economic, safety and regulatory requirements (Dutrieux et al., 2015a,b). Our objective is to find the optimal value of one parameter of the strategy, $x \in [-1; 0]$, so as to minimize the mean global cost of integrating about 20 megawatts of RES over 10 years.

The objective function is $f(x) = \mathsf{E}_S\left(C\left(x, S\right)\right)$, where S denotes a 10-year scenario (consisting of several time series, together with the characteristics of RES connection requests), E_S the expectation with respect to a random scenario, and $C(x, S)$ the cost of the strategy with parameter x applied to the scenario S. The computation of C is performed by an expensive-to-evaluate computer program. We assume evaluations of the form $\xi_i^{\mathrm{obs}} = C\left(X_i, S_i\right)$, where $S_1, S_2, \ldots$ are independent scenarios generated by the same scenario generator (and therefore identically distributed). This can be rewritten as $\xi_i^{\mathrm{obs}} = f(X_i) + \varepsilon_i$, where the variables $\varepsilon_i = C\left(X_i, S_i\right) - f\left(X_i\right)$ are independent and have zero mean. As shown in Figure 2 (left), the evaluation results are very noisy in this application. For the sake of simplicity, the noise variance is assumed to be a known constant (estimated based on a few result evaluations) and the variables ε_i, $i = 1, 2, \ldots$ will be assumed Gaussian.

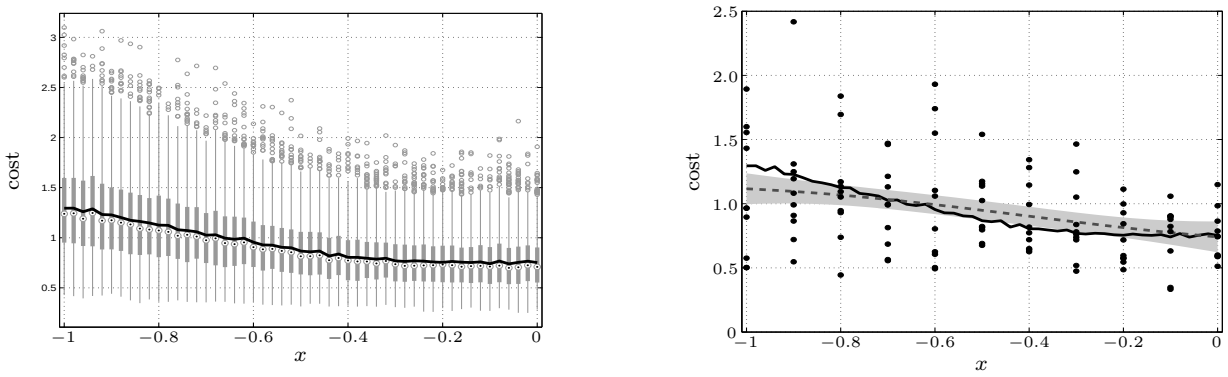

Figure 2: Left: Reference data. The search grid (on the x-axis) has $m = 51$ points. On each point, approx. 1000 evaluation results are available. The solid black line represents the empirical mean. Right: initial sample of 110 evaluations (11 batches of 10 evaluations). The dashed gray line represents the kriging mean. The grayed region represents pointwise credibility intervals with probability 95%.

We consider a budget of 2000 evaluations (without the initial sample) to find the minimizer among 51 candidate points linearly spaced in $[-1, 0]$. A batch of $K_0 = 10$ evaluations is performed at each iteration. We compare three ways of constructing $\underline{X}$: using the sampling criterion J'_n when $K = K_0 = 10$ (denoted as IAGO 10); using J'_n with $K = +\infty$ (denoted as IAGO $+\infty$); and, as reference, choosing X_{n+1} at random, uniformly in the set of candidate points (denoted as IID). The kriging model parameters are firstly estimated on an initial sample of 110 evaluations (11 batches of 10 evaluations as shown in Figure 2, right), then adjusted after each new batch of evaluations.

Figure 3 depicts the distribution of the estimated minimizer, the estimated minimum and the posterior entropy of the minimizer over the 500 optimization runs. IAGO $+\infty$ converges towards the area of interest faster than IID and IAGO 10. It is worth noting that a budget of 2000 evaluations does not suffice to locate the minimizer accurately. In fact, even 1000 evaluations at each candidate point (as in Figure 2, left), would not locate it much more precisely (result not shown).

4 Conclusion

We have proposed a new sampling criterion for the problem of global optimization in presence of very noisy evaluations, assuming that several evaluations are going to be made at a new evaluation point (even if they are not in practice). The proposed method has been applied to the optimization of a renewable energy integration strategy and shown to outperform plain IID sampling and the original IAGO criterion.

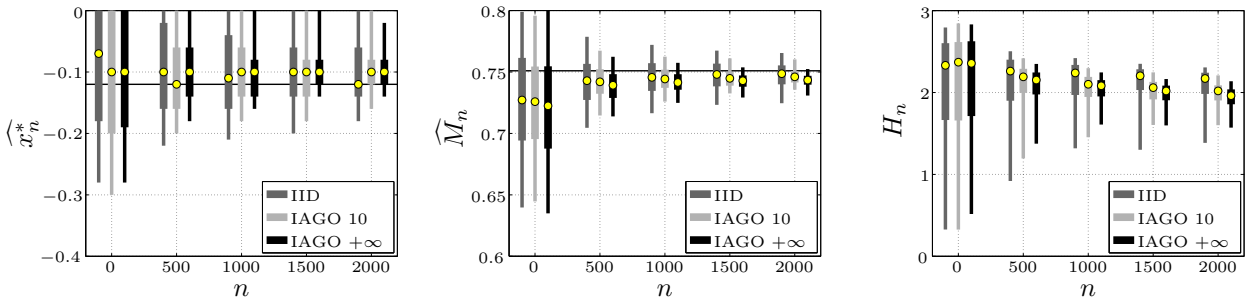

Figure 3: Distribution of the estimated minimizer $\widehat{x_n^*}$ (left), the estimated minimum $\widehat{M}_n$ (center) and the posterior entropy H_n of the minimizer (right) over 500 optimization runs. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers are the 5th and 95th percentiles. The thick black lines indicate the value obtained on our reference dataset.

5 Acknowledgements

This work was supported in part by the French National Research Agency (ANR) under PROGELEC 2012 grant program (APOTEOSE project, ref. ANR-12-PRGE-0012). The authors gratefully acknowledge its support.

References

J. Bect, E. Vazquez, et al. STK: a Small (Matlab/Octave) Toolbox for Kriging. Release 2.2.0. http://kriging.sourceforge.net, 2014.

H. Dutrieux, G. Delille, and B. Francois. Assessing the impacts of distribution grid planning rules on the integration of renewable energy sources (under review). Submitted to PowerTech 2015, June 29–July 2, Stockholm, 2015a.

H. Dutrieux, G. Delille, B. Francois, and G. Malarange. An innovative method to assess solutions for integrating renewable generation into distribution networks over multi-year horizons (under review). Submitted to CIRED 2015, June 18–15, Lyon, 2015b.

V. Picheny, D. Ginsbourger, and Y. Richet. Noisy expected improvement and on-line computation time allocation for the optimization of simulators with tunable fidelity. EngOpt 2010, Sept 6– 9, Lisbon, https://hal.archives-ouvertes.fr/hal-00489321v2, 2010.

V. Picheny, T. Wagner, and D. Ginsbourger. A benchmarck of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3):607–626, 2013.

E. Vazquez, J. Villemonteix, M. Sidorkiewicz, and E. Walter. Global optimization based on noisy evaluations: An empirical study of two statistical approaches. *J. Phys.: Conf. Ser.*, 135(012100):8pp, 2008.

J. Villemonteix, E. Vazquez, and E. Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509–534, 2009.